

Solution de l'exercice N°1 de la **Série N°2**

Exercice 1 Reprenons l'exercice 1 du TD 1.

On interroge 1587 étudiants de M2 sur la catégorie socioprofessionnelle de leurs parents. Les étudiants suivent différents cursus: écoles d'ingénieurs, écoles de commerce, universités scientifiques. Les résultats sont les suivants :

	Ouvriers	Employés	Cadres	Professions libérales
Écoles d'ingénieurs	50	280	120	20
Écoles de commerce	8	29	210	350
Universités Scientifiques	150	230	100	40

On veut étudier l'influence du milieu socioprofessionnel des parents sur le type d'étude des enfants. On rappelle que, dans l'exercice 1 du TD 1, nous avons déjà vérifié, par le test du Khi-deux qu'il existe une dépendance entre le milieu socioprofessionnel des parents et le type d'étude des enfants. On peut alors effectuer une analyse factorielle des correspondances sur les données.

1-Donner les centre de gravités, g_r et g_c , associés aux profils-linges et profils-colonnes.

2-Donner les matrices diagonales des profils-linges D_r et des profils-colonnes D_c .

3-Donner les matrices des profils-linges X_r et profils-colonnes X_c .

4-Calculer les deux matrices $A_r := X_r^t X_c^t$ et $A_c := X_c^t X_r^t$.

5-Calculer les valeurs propres et les vecteurs propres de A_r et de A_c . Que peut-on déduire?

6-Que représente g_r (resp. g_c) pour la matrice A_r (resp. A_c)?

Solution

Rappel: la matrice des effectifs observés est

$$N^* = \begin{pmatrix} 50 & 280 & 120 & 20 \\ 8 & 29 & 210 & 350 \\ 150 & 230 & 100 & 40 \end{pmatrix}.$$

La matrice des fréquences observées est

$$N = \begin{pmatrix} 50/1587 & 280/1587 & 120/1587 & 20/1587 \\ 8/1587 & 29/1587 & 210/1587 & 350/1587 \\ 150/1587 & 230/1587 & 100/1587 & 40/1587 \end{pmatrix}.$$

Les fréquences marginales des lignes sont

$$\begin{aligned} f_{1.} &= 50/1587 + 280/1587 + 120/1587 + 20/1587 = 0.29616 \\ f_{2.} &= 8/1587 + 29/1587 + 210/1587 + 350/1587 = 0.37618, \\ f_{3.} &= 150/1587 + 230/1587 + 100/1587 + 40/1587 = 0.32766 \end{aligned}$$

et les fréquences marginales des colonnes sont

$$\begin{aligned} f_{.1} &= 50/1587 + 8/1587 + 150/1587 = 0.13106 \\ f_{.2} &= 280/1587 + 29/1587 + 230/1587 = 0.33963 \\ f_{.3} &= 120/1587 + 210/1587 + 100/1587 = 0.27095 \\ f_{.4} &= 20/1587 + 350/1587 + 40/1587 = 0.25835 \end{aligned}$$

1) Le centre de gravité des profils-lignes:

$$g_r = (0.13106, 0.33963, 0.27095, 0.25835)^t.$$

Le centre de gravité des profils-colonnes:

$$g_c = (0.29616, 0.37618, 0.32766)^t.$$

Matrice diagonale des profils-lignes

$$D_r = \begin{pmatrix} 0.29616 & 0 & 0 \\ 0 & 0.37618 & 0 \\ 0 & 0 & 0.32766 \end{pmatrix}.$$

2) Matrice diagonale des profils-colonnes

$$D_c = \begin{pmatrix} 0.13106 & 0 & 0 & 0 \\ 0 & 0.33963 & 0 & 0 \\ 0 & 0 & 0.27095 & 0 \\ 0 & 0 & 0 & 0.25835 \end{pmatrix}.$$

3) Matrices des profils-lignes

$$\begin{aligned} X_r &= D_r^{-1}N = \begin{pmatrix} 0.29616 & 0 & 0 \\ 0 & 0.37618 & 0 \\ 0 & 0 & 0.32766 \end{pmatrix}^{-1} \\ &\times \begin{pmatrix} 50/1587 & 280/1587 & 120/1587 & 20/1587 \\ 8/1587 & 29/1587 & 210/1587 & 350/1587 \\ 150/1587 & 230/1587 & 100/1587 & 40/1587 \end{pmatrix} \end{aligned}$$

Donc

$$X_r = \begin{pmatrix} 0.10638 & 0.59574 & 0.25532 & 4.2553 \times 10^{-2} \\ 0.0134 & 4.8576 \times 10^{-2} & 0.35176 & 0.58627 \\ 0.28846 & 0.44231 & 0.19231 & 7.6924 \times 10^{-2} \end{pmatrix}.$$

Matrices des profils-colonnes

$$\begin{aligned} X_c &= D_c^{-1}N^t = \begin{pmatrix} 0.13106 & 0 & 0 & 0 \\ 0 & 0.33963 & 0 & 0 \\ 0 & 0 & 0.27095 & 0 \\ 0 & 0 & 0 & 0.25835 \end{pmatrix}^{-1} \\ &\times \begin{pmatrix} 50/1587 & 280/1587 & 120/1587 & 20/1587 \\ 8/1587 & 29/1587 & 210/1587 & 350/1587 \\ 150/1587 & 230/1587 & 100/1587 & 40/1587 \end{pmatrix}^T. \end{aligned}$$

Donc

$$X_c = \begin{pmatrix} 0.24039 & 3.8463 \times 10^{-2} & 0.72118 \\ 0.51949 & 5.3804 \times 10^{-2} & 0.42672 \\ 0.27907 & 0.48837 & 0.23256 \\ 0.04878 & 0.85366 & 9.7561 \times 10^{-2} \end{pmatrix}.$$

4) Calcul de la matrice $A_r := X_r^t X_c^t$:

$$A_r = \begin{pmatrix} 0.10638 & 0.59574 & 0.25532 & 4.2553 \times 10^{-2} \\ 0.0134 & 4.8576 \times 10^{-2} & 0.35176 & 0.58627 \\ 0.28846 & 0.44231 & 0.19231 & 7.6924 \times 10^{-2} \end{pmatrix}^T \\ \times \begin{pmatrix} 0.24039 & 3.8463 \times 10^{-2} & 0.72118 \\ 0.51949 & 5.3804 \times 10^{-2} & 0.42672 \\ 0.27907 & 0.48837 & 0.23256 \\ 0.04878 & 0.85366 & 9.7561 \times 10^{-2} \end{pmatrix}^T.$$

Donc:

$$A_r = \begin{pmatrix} 0.23412 & 0.17908 & 0.10332 & 4.4771 \times 10^{-2} \\ 0.46406 & 0.50084 & 0.29284 & 0.11368 \\ 0.21360 & 0.23362 & 0.28776 & 0.33150 \\ 8.8255 \times 10^{-2} & 8.6475 \times 10^{-2} & 0.31608 & 0.51006 \end{pmatrix}.$$

Calcul de la matrice $A_c := X_c^t X_r^t$:

$$A_c = \begin{pmatrix} 0.24039 & 3.8463 \times 10^{-2} & 0.72118 \\ 0.51949 & 5.3804 \times 10^{-2} & 0.42672 \\ 0.27907 & 0.48837 & 0.23256 \\ 0.04878 & 0.85366 & 9.7561 \times 10^{-2} \end{pmatrix}^T \\ \times \begin{pmatrix} 0.10638 & 0.59574 & 0.25532 & 4.2553 \times 10^{-2} \\ 0.0134 & 4.8576 \times 10^{-2} & 0.35176 & 0.58627 \\ 0.28846 & 0.44231 & 0.19231 & 7.6924 \times 10^{-2} \end{pmatrix}^T.$$

Donc

$$A_c = \begin{pmatrix} 0.40838 & 0.15522 & 0.35654 \\ 0.19716 & 0.67539 & 0.19448 \\ 0.39446 & 0.16939 & 0.449 \end{pmatrix}.$$

5) Calcul des valeurs propres et les vecteurs propres de A_r :

$$A_r = \begin{pmatrix} 0.23412 & 0.17908 & 0.10332 & 4.4771 \times 10^{-2} \\ 0.46406 & 0.50084 & 0.29284 & 0.11368 \\ 0.21360 & 0.23362 & 0.28776 & 0.33150 \\ 8.8255 \times 10^{-2} & 8.6475 \times 10^{-2} & 0.31608 & 0.51006 \end{pmatrix}.$$

Les valeurs propres

$$\lambda_1 = 1.0, \lambda_2 = 0.47966, \lambda_3 = 5.3109 \times 10^{-2}, \lambda_4 = 5.997 \times 10^{-7}.$$

Les vecteurs propres

$$u_1 = \begin{pmatrix} 0.25099 \\ 0.65040 \\ 0.51887 \\ 0.49474 \end{pmatrix}, u_2 = \begin{pmatrix} 0.25630 \\ 0.63057 \\ -0.17567 \\ -0.71122 \end{pmatrix}$$

$$u_3 = \begin{pmatrix} 0.723\,78 \\ -0.625\,73 \\ -0.248\,77 \\ 0.150\,7 \end{pmatrix}, \quad u_4 = \begin{pmatrix} 4.257\,9 \times 10^{-2} \\ -0.409\,29 \\ 0.801\,19 \\ -0.434\,47 \end{pmatrix}.$$

Calcul des valeurs propres et les vecteurs propres de A_c :

$$A_c = \begin{pmatrix} 0.408\,38 & 0.155\,22 & 0.356\,54 \\ 0.197\,16 & 0.675\,39 & 0.194\,48 \\ 0.394\,46 & 0.169\,39 & 0.449 \end{pmatrix}.$$

Les valeurs propres

$$\lambda_1 = 1.0, \lambda_2 = 0.479\,66, \lambda_3 = 5.310\,9 \times 10^{-2}.$$

Les vecteurs propres

$$u_1 = \begin{pmatrix} 0.510\,48 \\ 0.648\,41 \\ 0.564\,78 \end{pmatrix}, \quad u_2 = \begin{pmatrix} 0.383\,99 \\ -0.816\,03 \\ 0.432\,02 \end{pmatrix}, \quad u_3 = \begin{pmatrix} 0.709\,33 \\ -4.450\,4 \times 10^{-3} \\ -0.704\,86 \end{pmatrix}.$$

6) On remarque que l'ensemble des valeurs propres de A_c est inclus dans l'ensemble des valeurs propres de A_r .

7) Nous avons

$$\begin{aligned} g_r &= (0.131\,06, 0.339\,63, 0.270\,95, 0.258\,35)^T \\ &= \begin{pmatrix} 0.131\,06 \\ 0.339\,63 \\ 0.270\,95 \\ 0.258\,35 \end{pmatrix}. \end{aligned}$$

Remarquons que

$$1.915\,1g_r = 1.915\,1 \begin{pmatrix} 0.131\,06 \\ 0.339\,63 \\ 0.270\,95 \\ 0.258\,35 \end{pmatrix} = \begin{pmatrix} 0.250\,99 \\ 0.650\,40 \\ 0.518\,87 \\ 0.494\,74 \end{pmatrix} = u_1.$$

Comme u_1 est un vecteur propre de A_r associé à la valeur propre $\lambda_1 = 1$, donc g_r est aussi un vecteur propre de A_r associé à la même valeur propre $\lambda_1 = 1$. Avec le même raisonnement, on en déduit que g_c est aussi un vecteur propre de A_c associé à la même valeur propre $\lambda_1 = 1$.